How not to Characterize a Hard Choice

Kevin Reuter and Michael Messerli *

People are often faced with so called hard choices – also known as hard cases of comparison. In trying to characterize these hard choices, philosophers have made two central claims. First, failure of transitivity underlies hard cases of comparison. Second, using a random procedure is considered inappropriate in order to arrive at a decision in hard cases. While having some argumentative support, both claims primarily rely on expert intuitions. The results of the experiments we present in this paper challenge both claims, as well as the representativeness of expert intuitions that support these claims, by showing that most people (i) violate transitivity only if a hard choice is important, and (ii) find it appropriate to use a random procedure even in hard cases of comparison.

Keywords: Hard Cases; Small-Improvement Argument; Failure of Transitivity; Importance of Decisions; Random Choice; Rationality; Empirical Studies.

1 Introduction

Your favorite band is in town performing their best hits in a local club tonight. But it is also the last day of the season in which your home team plays against its main contender for the league title. Both events happen at the same time and you have tickets for both of them. Obviously, you cannot attend both events. Making a decision in such a situation is tough because one option seems to be neither better, nor worse compared to the

*This work is fully collaborative. Please address any correspondence to kevin.reuter@philo.unibe.ch and michael.messerli@philo.unibe.ch. We would like to thank an anonymous reviewer for this journal for extensive comments on the paper, as well as the audiences at conferences in Bern, Düsseldorf and Reading where this work was presented.

other, even when a small incentive like a 5% discount is added to one of the two options. These so-called hard cases of comparisons have attracted a lot of attention among philosophers (Boot 2009, Chang 2002a, Raz 1986) for they pose particularly fascinating problems for questions of comparability and rationality. However, before we can tackle these problems, it is crucial that an accurate description of hard choices is provided which is empirically adequate, i.e. that the features we use to characterize hard choices indeed single out these but not fictitious cases. This essay empirically investigates two particularly central claims, thereby raising serious doubts about each of them.

First, philosophers have claimed that in hard cases of comparison, people reject the application of a random procedure to arrive at a decision, i.e. it would be considered inappropriate to toss a coin in order to determine whether one should go to the concert or to the sports event. In contrast, when two options are perceived to be equally good, e.g. a choice between two cookies of the same type, people will find it appropriate to choose one option based on a random procedure. Importantly, we need to clearly distinguish the descriptive from the normative claim about using a random process. Whereas Schaber (1994, p. 158) has argued that we do find it inappropriate to toss a coin in hard cases, Chang (2002a, p. 684-685, 2012, p. 118) holds that it *is* inappropriate and thus that we *should* not toss in coin in such cases. In contrast, when people are faced with a choice in which both options are considered equally good, no such restrictions on the decision-making process are put forward by Schaber and Chang. Section 3 examines whether people think it inappropriate to toss a coin in hard cases, and how the results bear on both the descriptive and normative claims made in the literature.

The second claim that we are going to investigate states that hard cases of comparison manifest a failure of transitive reasoning, see e.g. Boot (2009) and Chang (2002a). In order to see why the violation of transitivity has received a lot of attention in debates on hard choices, let us consider the *small-improvement argument*, hereafter SIA: If a person asserts:

(1) x is neither better nor worse than y; E.g., a career as a lawyer is neither better nor worse than a career as an artist.

and:

(2) $x+$ is neither better nor worse than y; E.g., a career as a lawyer plus two additional holidays is neither better nor worse than a career as an artist.

and:

(3) $x+$ is better than x; E.g., a career as a lawyer plus two additional holidays is better than a career as a lawyer.

then we can logically conclude that x is neither better than, nor worse than, nor equally good to y.[1] Whether or not the conclusion of the small-improvent argument means that options in hard cases are incomparable (Raz 1986), on par (Chang 2002a), or roughly comparable (Parfit 1986) is an interesting question that is beyond the scope of this paper. However, while the question of the proper interpretation of the SIA has been hotly contested, most philosophers agree that the three premises of the SIA additionally reveal a failure of transitivity. If people were to reason transitively, then $(x+)$

[1] It is usually claimed in philosophical debates that additional premises, such as transitivity, are necessary to arrive at this conclusion (e.g. Espinoza 2008, p. 130; Gustafsson & Espinoza 2010, p. 755). However, the rules of first-order logic suffice to show that $\neg(x{=}y)$:

1	$\neg(x+ \succ y)$	premise 2 of the SIA
2	$x+ \succ x$	premise 3 of the SIA
3	$x = y$	assumption
4	$x+ \succ y$	2,3
5	$x+ \succ y \wedge \neg(x+ \succ y)$	1,4 $\wedge I$
C	$[\therefore] \neg(x = y)$	3,5 *reductio*

Assuming that x and y are equally good, we derived a contradiction – substituting x and y in premise (2). This leads to a reductio of assumption (3). The negation must be true, i.e. x and y are *not* equally good. Note that this reductio argument is conditional on premise 1 and 2. Thus, *if* premise 1 and 2 are correct - which people usually assume in this context - then it logically follows that x and y are *not* equally good.

should not be considered better than (x) when they also agree to (1) and (2) (see the next section for a formal analysis). Thus, if the three premises of the small-improvement argument hold for hard cases, as argued for by Chang (2002a), Boot (2009) and others, then people violate transitivity when being confronted with such choices. We have recently tested people's response profiles to two hard cases of comparison and found that a slight majority of the participants indeed failed to reason transitively.[2] Nonetheless, these results seem to call for further investigation for at least two reasons: (i) A substantial minority (around 40% of the tested subjects) did not violate the transitivity condition, which requires a more elaborate explanation. (ii) Only a specific subset of hard cases were presented to participants, namely, scenarios that involve important decisions. However, a wide range of examples of putative hard cases of comparison have been proposed in the philosophical literature, e.g. choosing between tea and coffee, going to one event or the other, deciding between a career as a lawyer and a career as a musician (Andreou 2011, Boot 2007, Chang 2002a, Chang 2012, Raz 1986). These and other decisions differ massively in the importance they have for people's lives. An alternative explanation of violations of transitive reasoning considers the interplay between hard choices and the importance these choices have for people's lives. Thus, people may well reason intransitively only when a hard choice is important.

But how can we understand the distinction between an important and an unimportant decision? An important decision has far reaching consequences. It enables and disables a whole series of other possibilities. Choosing between a career as a lawyer and a career as an artist, for example, is rather important because becoming a lawyer will foreclose many opportunities that an artist will be offered (and vice versa). In contrast, a consumption choice in a restaurant is rather unimportant because it does not enable and disable a host of other possibilities. People simply have a cup of coffee and, in most circumstances, move on with their lives in the same way as if they had taken

[2] See Messerli & Reuter 2016. Psychologists have conducted several experiments to show that people do not always have transitive preferences (e.g. Tversky 1969, Fishburn 1991, Loomes & Taylor 1992). Tversky (1969) for instance, argues that in a specific context like a multidimensional decision situation, people do violate so called *weak probabilistic transitivity*, i.e. people prefer x to y and y to z with a likelihood greater than 50% but do not prefer x to z in more than 50% of times. In this article, we do not adress the classic transitivity requirement, but cases of negative transitivity (see section 2.1). As far as we know, there is no empirical research so far, showing that people violate negative transitivity. We plan to extend our research to investigate violations of negative transitivity in probabilistic settings.

a cup of tea. In other words, the costs of correcting an important decision are much higher than an unimportant one. Becoming aware that studying law is not quite the right thing to do in one's life, is a very costly realization. In this paper, we will therefore put a particular focus on how the degree of importance might matter for the claims that have been put forward about hard cases, i.e. they manifest (i) failure of transitivity, and (ii) inappropriateness of choosing randomly.

Both putative characteristics of hard cases of comparison will be critically examined in this essay. To do so, we conducted six empirical studies. The results of the experiments we present in this paper show that: (*a*) If the decision process lacks importance, a substantial majority of the participants do not violate transitivity. Only if the decision is important, a slight majority fails the transitivity requirement. (*b*) The perceived inappropriateness of selecting randomly does not seem to be a characteristic that can be used to single out hard cases of comparison. When comparing the percentages of people who are willing to "let the coin decide", no differences were found between hard cases and those in which both options were considered equally good.

The paper will proceed as follows: In Section 2 we investigate whether people violate transitivity in a set of hard choices. Section 3 addresses the question of whether rejecting random selection is a successful test to identify hard cases. Section 4 concludes with a General Discussion. The Appendix presents the vignettes of the empirical studies.

2 Is Transitivity Violated in Hard Cases?

2.1 Theoretical Background

While there does not exist a consensus among philosophers on how to interpret hard cases, it is generally accepted that they characteristically involve a failure of transitivity. To see why, consider the following argument involving the premises of the small-improvement argument (SIA):

i. Failure of transitivity is constitutive of the premises of the SIA.

ii. The premises of the SIA underly hard cases of comparison.

∴ Failure of transitivity underlies hard cases of comparison.

In the introduction we have discussed the choice between a career as a lawyer and a career as an artist as an example of the SIA. In order to see more clearly how failure of transitivity is constitutive of the premises of the SIA, let us formalize the premises of the SIA:

1 $\neg(x \succ y) \wedge \neg(y \succ x)$	premise
2 $\neg(x+ \succ y) \wedge \neg(y \succ x+)$	premise
3 $x+ \succ x$	premise

'x', 'y' and '$x+$' stand for the options in decision-making situations.[3] The formalized premises of the SIA violate the following kind of transitivity:

$$(x+ \succ x \wedge \neg(y \succ x)) \rightarrow x+ \succ y$$

Thus, if people endorse all three premises of the SIA, they reason intransitively.

It is important to notice that the transitivity under consideration is not of the form such as x is better than y, y is better than z, thus x is better than z. Instead, the transitivity is logically equivalent with following negative transitivity principle:[4]

$$\neg(x+ \succ y) \wedge \neg(y \succ x)) \rightarrow \neg(x+ \succ x)$$

While the first premise (i) of the argument stated above is therefore logically true, the second premise (ii) seems to allow for empirical falsification. It states that all three premises of the SIA are endorsed in hard cases of comparison. Several philosophers have argued for the truth of premise (ii). Chang, for instance argues that the SIA holds for all kinds of hard cases, such as a choice between a cup of tea and a cup of coffee or a choice between two careers (Chang 2002b, p. 126). Thus, in all these cases an agent makes the following judgments: (1) x is neither better nor worse than y, (2) $x+$ is neither better nor worse than y, (3) $x+$ is better than x. Given the argument stated above, Chang commits herself to the view that in all these hard choices an agent violates transitivity. A similar position is held by Boot (2009, p. 89).

[3] Read '$\succ$' as 'is better than'. Read '$\prec$' as 'is worse than'. Read '$=$' as 'equally good'.

[4] For example, Bridges & Mehta reflect on negative transitivity (Bridges & Mehta 2013).

Some philosophers might not agree that (ii) is an empirical premise but is rather part of the definition of what it is for a choice to be hard. Accordingly, if a person does not also endorse premise (3) of the SIA then the choice she faces is not hard for her. We disagree with this conception of a hard case for reasons that will become manifest during the course of this paper. We will, however, take up this issue in the General Discussion when we present several arguments in favor of the claim that failure of transitive reasoning is not constitutive of hard cases of comparison.

How do we propose to separate hard choices from choices that are not hard? Hard choices seem to be different from choices in which options are considered equally good – hereafter simple cases – in that a small-incentive does not make any difference in the decision process. Thus, for those people who endorse premise (1) and premise (2) of the SIA, the choice seems to be hard, because not only is neither option better or worse than the other, it is also the case that an improved option is not considered better than the other option. In contrast, when the original options are considered neither better nor worse, but a small improvement makes a difference in the perception of the options, then the original options seem to have been considered of equal value, which we call simple cases.

So, do people really violate transitivity in hard cases? While it seems intuitively correct to say that people prefer a slightly improved option ($x+$) compared to an unimproved choice (x), empirical evidence has not yet been collected in support of this intuition. Study 1A and 1B were designed to test people's response patterns in hard cases to examine whether people indeed reason intransitively.

2.2 Study 1A

Methods

In order to investigate (a) whether people violate transitivity in hard cases of comparison, and (b) how the importance of the decision influences people's responses in these cases, we selected three cases of varying importance: (a) a low importance decision between two drinks; a choice of medium importance between two events; and a high importance decision between two courses of study. 151 participants were randomly assigned to one of the three cases. All participants were recruited through Amazon's Mechanical Turk and reimbursed for their participation. The 'Study scenario' read as follows (all vignettes are attached in the Appendix):

Some people think that studying art is neither better nor worse than studying law. Others believe that studying philosophy is neither better nor worse than studying economics.

Imagine you have got the choice only between studying two different courses. Please name two courses for which YOU could not tell which one of those is better or worse than the other for studying.

Participants then had to write down two courses of study which they thought fulfilled the condition of being neither better nor worse. At the next stage, we identified those participants for whom a small-improvement would not make any difference in the decision process. All those participants who chose a slightly improved option $(x+)$ over (y), would indicate that they considered the two original options (x) and (y) to be equally good. The other participants seemed to be indeed faced with a hard case of comparison. In order to compare the three cases, we used an incentive that had the same relative value for all three decisions – a 5% discount on one of the options. Thus, the second vignette presented to the participants read:

Now, imagine that the university tells you that you will get a discount of 5% on the costs of one of these courses of study. Keeping in mind the two courses you selected before, how would you evaluate your situation?

- The course plus a small discount is better than the other course without a discount.
- The course plus a small discount is worse than the other course without a discount.
- The course plus a small discount is neither better nor worse than the other course without a discount.

In all three scenarios (drink choice, event choice, study choice) we also asked the participants a third question, designed to investigate whether people indeed violate transitivity in hard cases of comparison:

Now, lastly, imagine you have got the choice between the course of study for which you were offered a 5% discount and the very same course without getting the discount. How would you evaluate your situation?

- The course plus a small discount is better than the very same course without a discount.
- The course plus a small discount is worse than the very same course without a discount.

- The course plus a small discount is neither better nor worse than the very same course without a discount.

Again, the participants were requested to assent to one of the three propositions above.

Results

We started the survey asking people to name two drinks, two events, or two courses for which they could not decide which one of them is better or worse than the other. In the second vignette, we offered a 5% discount to one of the two options. 38% of the participants in the drink case, 37% of the participants in the event case, and 51% of the participants in the study case, indicated that this was a hard choice by selecting the 'neither better nor worse' option. The third question investigated the transitivity requirement. Of the 21 participants for whom the drink choice was a hard case, 6 participants decided that the improved option $(x+)$ was better than (x), whereas 15 decided that neither option was better than the other. A similar result obtained in the event choice with 3 out of 14 participants opting for the 'better' option and 11 participants for the 'neither better nor worse option'. Only in the study case did a majority believe that $(x+)$ was better (17 participants out of 29) than (x) compared to 12 participants who thought $(x+)$ was neither better nor worse than (x). The percentages of all three experiments are shown in Figure 1. No participant in any of the three cases thought that $(x+)$ was worse than (x).

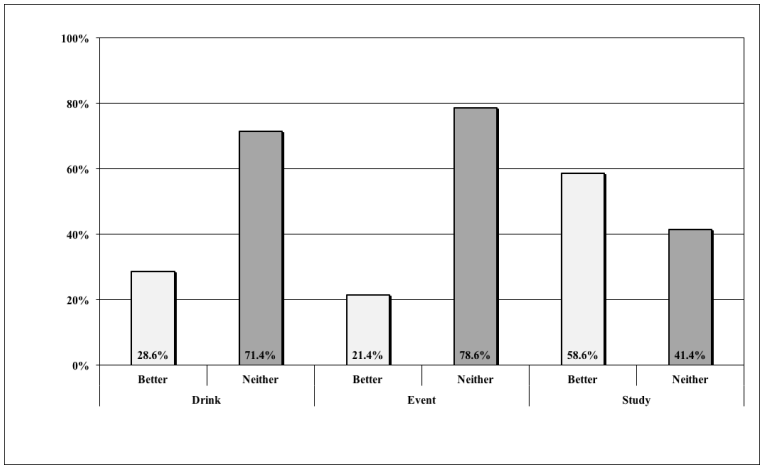

Figure 1 Response percentages to the third question (testing for violations of transitivity) comparing a slighly improved option $(x+)$ with the original option (x). We excluded the 'worse' option in this figure, because it was not selected by any of the participants.

We conducted Pearson's Chi Square tests to investigate whether data from any of the three experiments yielded significant differences from the 50% mark – only taking into account 'better' and 'neither better nor worse' responses. None of the tests were significant. Drink Case: χ^2 =2.12, $p = 0.145$; Event Case: χ^2 =2.49, $p = 0.115$; Study Case: $\chi^2 = 0.43$, $p = 0.510$. We then compared the scenarios with each other using conservative Bonferroni correction. Both the Drink case and the Event Case differed significantly from the Study Case: χ^2 =4.43, $p = 0.035$ (Drink vs. Study) & χ^2 =5.25, $p = 0.022$ (Event vs. Study), supporting the view that the importance of the decision had an effect on the transitivity question.[5]

2.3 Study 1B

Methods

Study 1A indicates that a majority of people violate transitivity in hard choices only if the decision is sufficiently important. However, this conclusion can be challenged: the three scenarios not only vary in terms of importance, but vary on a host of other dimensions, e.g. factual uncertainty, kind of activity, amount of money, length of time spent doing the activity, etc. In order to focus more narrowly on the importance of the decision, we designed a second study in which the vignettes are not only more easily comparable but in which we also only varied one factor. The vignettes read:

Scenario 1

Some people think that living in Australia is neither better nor worse than living in Sweden. Others believe that living in Switzerland is neither better nor worse than living in Ireland. Imagine you have got the choice between spending the **next year / next 25 years** living in one of two different countries. Please name two countries (other than the

[5] In order to control for a possible order effect on the transitivity question, we conducted a study in which we asked 103 participants a single question about whether a slightly improved option ($x+$) is considered better, worse, or neither better nor worse than an unimproved option (x). The participants were randomly assigned to the drink choice, event choice, or study choice. In all three cases, an overwhelming majority considered option ($x+$) to be better than option (x): 85.3% in the Drink case, 91.1% in the Event case, and 88.6% in the Study case. The percentages of the responses were in all three cases significantly different compared to the responses in the original study: Drink Case: χ^2 =18.05, $p < 0.001$; Event Case: $\chi^2 = 25.83$, $p < 0.001$; Study Case: χ^2= 8.03, p = 0.005. This demonstrates that the context in which the decision is made has a particularly strong influence on people's choices (see Discussion 2.4. for further elaboration of this point).

United States) for which YOU could not tell which one of those is better or worse than the other to live in.

People were prompted to name two countries, Country A and Country B, which would then play a significant role in the decision processes in Scenario 2 and Scenario 3.

Scenario 2

Now, imagine you have got the choice between spending the **next year / next 25 years** either living in country A plus receiving an extra dollar per week or living in country B and not receiving an extra dollar per week. How would you evaluate your situation?

- Living in country A for one year / 25 years plus receiving an extra dollar per week is **better** than living in country B for one year / 25 years.
- Living in country A for one year / 25 years plus receiving an extra dollar per week is **worse** than living in country B for one year / 25 years.
- Living in country A for one year / 25 years plus receiving an extra dollar per week is **neither better nor worse** than living in country B for one year / 25 years.

Scenario 3

Now, imagine you have got the choice between spending the **next year / next 25 years** either living in country A plus receiving an extra dollar per week or living in the same country A and not receiving an extra dollar per week. How would you evaluate your situation?

- Living in country A for one year / 25 years plus receiving an extra dollar per week is **better** than living in country A for one year / 25 years.
- Living in country A for one year / 25 years plus receiving an extra dollar per week is **worse** than living in country A for one year / 25 years.
- Living in country A for one year / 25 years plus receiving an extra dollar per week is **neither better nor worse** than living in country A for one year / 25 years.

As can be seen from the three scenarios, we only varied information about the length of the stay abroad, i.e. whether the person would spend one year or 25 years in a different country. Obviously, a decision as to stay one year abroad is much less important than a stay for 25 years.

Results

161 participants were recruited through Amazon's Mechanical Turk and randomly assigned to either the *one year* condition (N = 83) or the *25 years*

condition (N = 78). Of the 83 participants in the *one year* condition, 39 subjects selected 'neither better nor worse' when a small incentive, i.e. an extra dollar per week, was added to the first option, thereby indicating that the choice was indeed hard for them. 13 out of the 39 participants (33%) violated the transitivity condition by selecting the option 'better' in the third scenario when comparing $x+$ with x. 67% (26 participants) did not violate transitivity. In the *25 years* condition, 42 participants selected 'neither better nor worse' when a small incentive was added. A greater percentage, i.e. 52% (N = 22) of the participants failed the transitivity condition whereas 48% selected 'neither better nor worse'. A Pearson's Chi Square test revealed a marginally significant difference between the "unimportant" and the "important" condition: χ^2 =2.99, p = 0.084.

2.4 Discussion

Several conclusions can be drawn from the results of Study 1A and Study 1B. First, when people deliberate on hard choices, e.g. between going to a concert or attending a sports game, it does not follow automatically that people violate a crucial transitivity requirement that has been singled out as a characteristic of hard cases of comparison. In two (drink choice, event choice) out of three cases in Study 1A, a majority of the participants did not prefer an improved option $(x+)$ over the original (x) when having previously declared not to favor either (x) over (y), and $(x+)$ over (y). Similarly, in Study 1B, a majority of the participants who were confronted with a choice between going to one of two countries for one year, failed to violate the transitivity condition.

Second, the importance of the scenarios seemed to play an important role in people's reasoning. In the study case, a majority violated transitivity by favoring a discounted course of study compared to the non-discounted course. Differences between the study case and the other two cases turned out to be significant. As we have mentioned above, it might be objected that the differences between the study case and the other two scenarios should not be explained by the importance of the decision but was rather due to other possible factors that were varied in Study 1A.[6] In order to address this objection we conducted Study 1B in which only the time of stay in

[6] We would like to thank a reviewer of this journal for stimulating a discussion on this issue.

one of two countries was varied. In the more important decision (25 years) a greater number of participants violated transitivity than in the less important condition (one year). One might respond that it was rather the length of time that mattered in people's decision process not so much the importance attached. However, we cannot see any plausible reason why people should reveal different response patterns given the length of time independently of the importance attached to it.

Third, people's responses to the transitivity question were highly context-dependent. An overwhelming majority prefered an improved option ($x+$) over the original one (x) when asked independently of a hard case (see footnote 5). However, if people were confronted with a hard choice involving a decision between ($x+$) and (x), then most people did not seem to prefer ($x+$) at least when the decision was not very important. In the General Discussion we discuss how the results of Study 1A and Study 1B can be explained given that they seem to contradict the widely held claim about failure of transitivity as a mark of hard cases.

3 Do people reject the toss of a coin in hard cases?

3.1 Theoretical Background

If you are confronted with a choice between two equally appealing law firms that only differ in regards to their names, or with a choice between two cookies of the same type, flipping a coin seems to be an adequate strategy.[7] Do people find it equally appropriate to toss a coin in a hard case? Or is it characteristic of a hard case, in contrast to a case in which both options are considered equally good, that a random choice is perceived to be inappropriate? Some philosophers have claimed that in hard cases of comparison, we do not find

[7] One could also say that picking is an adequate strategy. Chang for instance writes: "When the alternatives are equally good, you pick rather than choose because the reasons don't favor one option over the other but give you license to select arbitrarily." (Chang 2012, p. 117). Where options are equal, it is not worhwhile to go through the process of choosing. Picking makes perfect sense (Ulmann-Margalit & Morgenbesser 1977). There is a difference between picking an option and using a stochastic process, such as tossing a coin. Unconscious biases may influence how we pick one option over the other. However, for sake of simplicity let us regard both picking and flipping a coin as a random choice.

it appropriate to toss a coin in order to arrive at a decision. E.g., Schaber (1994) writes:

> *"In a choice between two exactly equally valuable options, we would be willing without hesitation, to toss a coin. This does not apply to two options which are roughly equal. We do consider it inappropriate to toss a coin in such a situation."*[8] (Schaber 1994, p. 158)

Schaber's suggestion is interesting and, perhaps, prima facie plausible. However, he does not provide any independent evidence for his proposal but seems to rely solely on his own intuitions about easy and hard cases of comparison. Other philosophers have voiced similar claims, e.g. Chang states:

> *"The resolution of perplexity in hard cases is different in nature. Suppose Jack puts the ordered pair (philosophical career x, legal career y) in the 'better than with respect to goodness as a career' pile. Whether the pair is to be sorted in this way is not, however, a matter of arbitrary stipulation, a coin flip cannot appropriately determine how the case is to be resolved."* (Chang 2002b, p. 138)

It should be noted that whereas Schaber focuses on the descriptive question of whether people find it inappropriate and hence whether people will toss a coin in hard cases, Chang, instead, makes a normative claim about the inappropriateness of using a coin toss, i.e. she asks whether people *should* toss a coin in hard cases. Chang's main argument for why she believes a coin toss is hardly ever an appropriate way to determine the right choice in a hard case, is a psychological one. She maintains that an agent needs to commit herself to her decision in hard cases.: "Commitments are by their nature willings of some kind." (Chang 2013b, p. 92). Accordingly, making a commitment means "willing something to be a reason" (Chang 2013b, p. 93), whereby she explains this as follows: "Willing something to be a reason is the activity of placing your will – your very agency – behind its being a reason." (Chang 2013b, p. 93). In other words, decisions in hard cases should be based on reasons but not on random selection, e.g. a coin toss:

[8] The German original reads: „Bei zwei genau gleich wertvollen Optionen wären wir ohne zu zögern bereit, eine Münze zu werfen. Das trifft auf ungefähr gleich wertvolle Optionen nicht zu. Bei solchen Optionen halten wir das Werfen einer Münze für unangemessen."

> *"Far from being cases in which you must abandon your rational agency, hard choices arguably present opportunities for exercising our rational capacities to their fullest. In thinking about which career to pursue, you might marshal all your cognitive and non-cognitive resources to settle on an alternative that best expresses who you are as a rational agent."* (Chang 2012, p. 118)

Following Chang, we have pyschological reasons to reject a random selection in hard cases of comparison. Other philosophers have put forward the claim that it is rational to randomize in hard cases. E.g., Peterson (2013) argues that an agent should decide randomly in hard choices. The idea is that an agent who does not flip a coin will be exploited in a modified *money pump situation*.[9]

So, what is the 'empirical reality'? Do people really reject a random procedure to arrive at a decision in hard cases? More specifically, are there any significant differences between hard choices and easy choices in terms of whether or not people think it appropriate to toss a coin? Does it matter whether the choice is important? In order to investigate these questions, we designed two experiments that would allow us to determine whether coin tossing is considered an inappropriate means to arrive at a decision in hard cases of comparison.

3.2 Study 2A

Methods

We used the same vignettes as in Study 1A, i.e. 146 participants were randomly assigned to either the drink case, the event case, or the study case. We first asked the participants to name two drinks (respectively, two events or two courses) for which they could not tell that one of the two options was better or worse than the other. We then presented participants with the small-improvement choice to separate those people who considered the choice to be hard from those who thought of the options as being of equal value (see Appendix for the exact wording of the vignettes).

[9] Let us briefly illustrate Peterson's point: Suppose an agent chooses y instead of $x+$, and x instead of y. Then the agent will pay a small amount of money to move from x to $x+$. If we play this game ad infinitum, the agent loses all his / her money. In contrast, an agent who decides randomly will not be exploited (Peterson 2003, p. 129).

The third and last vignette, did not, however, ask the transitivity question, but instead prompted the participants to state whether in the original choice, flipping a coin would be an appropriate means to arrive at a decision. The vignette in the Study case read:

> In the beginning, we asked you to imagine two courses of study for which you could not tell which one of those courses is better or worse than the other. Would flipping a coin be an appropriate way to decide in favor of one of the options?

- Yes
- No

All participants were recruited on Mechanical Turk and compensated for their participation.

Results

Based on the participants' responses to the second question, we divided the participants into two groups. Those for whom the options were a hard case of comparison, i.e., those who chose 'neither better nor worse' when a small discount was offered to one of the two options, and those participants for whom a small improvement made a difference, i.e., those who chose 'better' when a small discount was offered. As mentioned before, we will call this the 'simple case'. Table 1 lists the absolute numbers as well as the percentages of participants that opted in favor of tossing a coin.

	Absolute numbers		*Percentages*	
	Simple Case	Hard Case	Simple Case	Hard Case
Drink Scenario	21 (out of 33)	11 (out of 21)	64%	53%
Event Scenario	18 (out of 24)	11 (out of 14)	75%	79%
Study Scenario	14 (out of 25)	15 (out of 29)	56%	52%

Table 1: Positive answers to the 'tossing a coin' - question in absolute and relative numbers.

Pearson's χ^2 tests were carried out for all three scenarios to compare the numbers between the hard case and the simple case. None of the three scenarios yielded any significant differences between the two types of cases: Drink Scenario ($\chi^2 = 0.67$, p=0.412); Event Scenario ($\chi^2 = 0.06$, p=0.803); Study Scenario ($\chi^2 = 0.10$, p=0.754).

3.3 Study 2B

Methods

While the results of Study 2A suggest that the importance of a hard choice might influence whether people consider it inappropriate to toss a coin in making a decision, the data do not provide us with a clear picture. Furthermore, we already argued that various confounds can be identified that might have skewed the results. We therefore follow the same strategy as with the transitivity studies, and used the vignettes of Study 1B to test whether the importance of a decision matters for how appropriate people find it to toss a coin. The only difference between the two cases consisted in the length of the time that people would spend in either of two countries. We presented the same participants of Study 1B with the following vignette:

> In the beginning we asked you to imagine a choice between living in country A for **one year / 25 years** or in country B. Would flipping a coin be an appropriate way to decide in favor of one of the options?

People then had a simple choice between Yes and No.

Results

Of the 83 participants who were assigned to the less important decision of choosing to spend one year abroad, 62 participants (74.7%) considered it appropriate to toss a coin as a means to make a decision. In contrast, only 39 out of 78 participants found a coin toss appropriate in the more important decision on where to spend the next 25 years. A χ^2 test revealed that this difference was highly significant: $\chi^2 = 10.41$, p=0.001.

3.4 Discussion

The empirical studies that we carried out to investigate the inappropriateness condition as a characteristic feature of hard cases, delivered a clear verdict: Study 2A demonstrates that people seem to be just as willing to opt in favor of tossing a coin in hard cases as they are in situations in which they consider the choice to be simple. Study 2B shows that whether or not the use of a random procedure is considered appropriate partially depends on whether the choice is important or not.

Objection 1

It might be objected that the participants in our study did not understand what it means to flip a coin. In order to ensure that the participants were aware of the consequences of flipping a coin, we asked them to explain their decision in Study 2A. The responses strongly suggest a proper understanding of the appropriateness of the coin toss. Explanations did not differ between those participants for whom the choice was hard and those for which the choice was simple.

Examples of typical explanations in favor of coin tossing in hard cases:

> *"Since there is no better option, randomly choosing would suffice.", "Both fields are incredibly valuable to study; as such, I could not pick which one to take by myself without some random generation involved."*

Examples of explanations in favor of coin tossing when options are seen as a simple choice.

> *"If you cannot decide which is better or worse, flipping a coin is a neutral way to decide which to take", "Because its the exact same odds as me just picking one."*

Similarly, when the participants rejected such a coin toss, they seemed to understand the process clearly:

Examples of explanations against coin tossing in hard cases:

> *"Courses decide your future, you really can't be leaving to luck", "Can't leave a major decision to random chance."*

Examples of explanations against coin tossing when options were not seen as a hard choice:

> *"Randomly selecting your education is never a wise choice.", "I would rather take the course that I feel like I would do better at."*

Objection 2

A more serious objection challenges the correctness of the order of the different vignettes that we presented to the participants. Note that after prompting participants for options that they considered neither better nor worse, we did not immediately ask them about the rationale of a coin flip but rather presented them with the vignette in which one option was slightly improved. Thus, if the questions about flipping a coin had been asked straight after the first vignette, results might have been different. To test the possibility of an order effect, we therefore repeated Study 2A with one important difference. This time, we reversed the second vignette and the third vignette, such that the participants first answered whether a coin flip is appropriate and then had to face a choice between $(x+)$ and (y). The results show no significant differences between the two types of cases. Thus, the experiment provides further support that indeed people find coin tossing as appropriate in hard cases of comparison as in simple cases.[10]

4 General Discussion

In a series of experiments, we investigated two claims. First, it has been argued that transitivity is violated in hard cases of comparison. Second, it has been claimed that people find it inappropriate in hard choices to make a decision in favor of one of the two options based on a random procedure. In making these claims, philosophers have primarily relied on their own intuitions. However, expert intuitions may not reflect the behavior of people when they are in fact confronted with decisions in hard cases of comparison. Even though the results of the experiments raise serious worries for the claims that have been made by philosophers based on their intuitions, we cannot conclude from these results that those intuitions are mistaken. To put expert intuitions to the test, experiments need to be set up that put participants in situations in which they face a hard choice and in which their decisions have a direct impact on their future lives, e.g. by indeed getting a cup of coffee but no tea.[11] Instead, we asked laypeople to reflect on how they themselves

[10]Using Pearson's χ^2 test, we compared the results between the hard case condition and the simple case condition for all three scenarios. None of the tests yielded a significant difference between the two conditions (Drink: $\chi^2 = 0.74$, p=0.388, Event: $\chi^2 = 0.66$, p=0.416, Study: $\chi^2 = 0.01$, p=0.916).

[11]Note, that unimportant hard choices, like deciding between two drinks, can easily be realized in an experimental setting. Confronting participants with important hard choices

would have behaved in hard cases of comparison. Thus, despite the results we report in this paper, it might still be the case that the experts' intuitions are correct. However, while we do not believe that asking people about their preferences in hard cases will necessarily match their behavior precisely, collecting responses about people's preferences are much more likely to map their behavior than the intuitions of experts. Several empirical studies have shown high consistency between people's stated preferences and their revealed preferences when it comes to e.g., vaccination behavior (Lambooij et al. 2015), consumers' choices (Loureiro et al. 2003) and family planning (De Silva 1991).

Given that asking people about their own preferences in hard choices is a much more promising way to get at people's true behavior than relying on expert intuitions about other people's preferences, how do both claims fare in light of the empirical results we presented in this paper?

4.1 Inappropriateness of Coin Tossing

One of the claims under investigation was that using a random procedure is deemed inappropriate in hard choices. The results strongly suggest that this claim is mistaken. In all three scenarios of study 2A we found that a majority of the participants did not feel it would be inappropriate to toss a coin in order to arrive at a decision in a hard case of comparison. One might remark that many more people will use a coin toss or any other random means in easy cases of comparison, e.g. when deciding between two yoghurts of the same type or between two careers at two equally appealing companies. While we have not tested people's responses in these cases, we do not doubt that in easy cases more people will be happy to toss a coin or throw a dice. However, the claim under investigation was not that slightly fewer people would feel that it would be inappropriate to toss a coin but rather that *most* people would do so. However, no significant differences were found when comparing the numbers between those who considered either of the three cases hard, and those that did not.

While the results presented in this paper seem to directly contradict Schaber's descriptive position, what is the bearing of these results on the normative position advocated by Chang? Obviously, we cannot derive the normative

in an experimental setting, is extremely difficult both from a financial as well as from an ethical point of view.

claim that it is appropriate to toss a coin in hard cases from the descriptive claim that we think it is appropriate. However, we believe Chang should provide an explanation as to why a majority of people hold an apparently mistaken picture about the appropriateness of tossing a coin. Chang will presumably consider the situation as one in which the reasons for why people should not toss a coin in hard cases are not transparent to them, i.e. there are reasons for rejecting a coin toss in hard cases, but many people have not yet reflected on them. The most plausible argument in favour of Chang's position can be constructed from her writings (2012, 2015) on people's commitments and the need to identify themselves with their decisions. This argument has the following form:

a. People need to identify with their decisions in hard cases.[12]

b. If people need to identify with their decisions in hard cases, then it is inappropriate to toss a coin in order to arrive at a decision in hard cases.

∴ It is inappropriate to toss a coin in order to arrive at a decision in hard cases.

Let us look in greater detail into the plausibility of the first premise:

When it comes to identifying oneself with a decision, the more crucial factor might well lie in the importance of the decision for a person but not in whether the decision is hard or simple. Especially Study 2B suggests that fewer people are willing to toss a coin in important cases. Thus, when being confronted with a decision that has relatively little impact on people's future lives, most people do not seem to feel an urgency to identify with their decision even though the choice is hard. In contrast, a decision as to whether a person should study law or philosophy, or a decision where to spend the next 25 years, will partially determine how that person's personality will be shaped. Therefore, there might be a much greater need to put one's agency behind important decisions. Hence, the need to identify with one's decision seems to apply more strongly to important decisions regardless of whether the choice is hard or simple.

[12]This premise is not meant to specify a moral obligation to identify with a decision in hard cases, but merely claims that there is a certain psychological need for such an identification.

4.2 Failure of Transitivity

A very prominent role in debates on hard cases is also taken by the claim that people violate transitivity when facing hard choices. The results we presented in Section 2, however, show that in less important hard cases, a majority of the participants failed to violate transitivity by not prefering a slightly improved option ($x+$) over the original option (x). What at first glance appears very unusual – Why should people not choose the better option? – can be given a sensible interpretation by drawing on Hsieh's concept of clumpiness (Hsieh 2005). According to Hsieh's view, we often compare items using a so called covering consideration (a decision criterion) that is clumpy, i.e. we sort items into clumps depending on how well they fulfil the relevant respects that comprise a certain criterion. 'Goodness as an undergraduate student article' for instance, is a clumpy decision criterion: We might have exam papers worth an 'A', exam papers worth a 'B' and so on, whereby these clumps "(...) reflect the smallest units of measurement for purpose of comparison." (Hsieh 2005, p. 198). How exhaustive these clumps are depends on the purpose of the comparison. Thus, another grading system with a more fine-grained resolution may function in terms of 'A+' exam papers 'A' exam papers 'A-' exam papers etc. The important point made here is that two options can be equally good according to one and not equally good according to another comparison, or as Andreou puts it: "In Hsieh's view, two options O1 and O2, and two resolutions R1 and R2, can be such that A is better than B when resolution R1 is in play and A is equal (not just roughly equal) to B when resolution R2 is in play." (Andreou 2015, p. 15).

Applying these theoretical considerations to our problem at hand, one might argue that given the resolution of the comparison people have in mind, the incentive ('+') is not relevant to the comparison. Thus, we believe a promising explanation of the results can be provided by analyzing the importance the incentive played in people's decision process: Those participants who considered the decision to be hard, were not persuaded by a small incentive to prefer the improved option ($x+$) over (y). This suggests, that the amount of the incentive was insufficient to "escape" the clump that was defined by the original options. It is now likely that the perception of the incentive as insufficient *carried over* from the second question to the third question (transitivity question). Thus, although the incentive "should" have played a pivotal role in the transitivity question, people seemed to have thought in the same categories or clumps, and have maintained the same

attitude towards the incentive in both scenarios: Thus, they did not prefer the improved option.

However, whereas in the drink and event case around 75% of the participants did not violate transitivity, only 40% failed to violate transitivity in the study case. Thus, the carry-over effect did not dominate people's responses in the hard choice scenario that had the greatest importance for the decision maker. Any comprehensive explanation should account for these differences.

The following explanation seems to be forthcoming: The incentive of a 5% discount was chosen in order to ensure that all three cases remain comparable. However, the absolute value of the incentive differs between all three cases. While in the drink and event case, the absolute value of the incentive amounts to a few dollars, in the study case, the absolute value can easily reach a few hundred or even thousand dollars. Hence, awareness of the absolute value may have counteracted the perception of the incentive as insufficient – at least to a certain extent. An initial reaction to chose "neither better nor worse" in the study case may be overridden by a secondary reasoning process to the effect that a lot of money would be saved if the discounted study course were chosen.

If this explanation is along the right lines, then it would be premature to dismiss people's responses as a mere bias which do not reflect the real preferences of the participants. Instead, people's preferences seem to change in hard cases of comparison such that the transitivity requirement is not violated.

Before we conclude, let us consider one last objection, to which we have already alluded in Section 2, against the results we obtained in regards to the transitivity claim. Some philosophers define hard choices as cases in which one option is neither better, nor worse, nor equally good than the other. We have argued above that in order to logically conclude that both options are not equally good, people need to violate the transitivity condition. Thus, it might be argued that whether or not people violate transitivity in hard cases is not an empirical matter but rather definitional. Accordingly, our experiments do not show that people do not fail the transitivity condition in hard cases, the results instead reveal that these cases are simply not hard cases of comparison. However, this response seems unconvincing. First, restricting hard cases in the way suggested would mean that many decisions are hard only for a tiny fraction of the population. Most cases that have been cited and discussed in the literature would not be real-world cases but rather fictitious hard cases of comparison. Second, this move would only lead to a

verbal dispute. Choices in which people do not believe one of the options to be better or worse even if an incentive is added, continue to exist and represent fascinating objects of research. Not calling them "hard cases" only changes them in name but does not change their philosophical attraction. Thus, future research should leave open the possibility that in many hard cases of comparison people do not violate transitivity.[13]

5 Conclusion

Hard cases raise fundamental questions for rational choice. For the last thirty years, philosophers have discussed hard cases of comparison in rational decision-making processes. The philosophical consensus states that hard cases can be characterized by intransitive reasoning and the perceivced inappropriateness of using a random process to make a decision. Both claims are first and foremost supported by the intuitive plausibility of exemplary cases, i.e. philosophers have relied on their own intuitions to evaluate the truthfulness of these claims. We have therefore started to investigate the two claims empirically. Our results lead us to disagree with the philosophical consensus in two important respects: First, rejecting a random procedure is not a characteristic of a hard case. People find it appropriate to toss a coin in order to arrive at a decision. Second, failure of transitivity seems to be an integral part of hard choices only insofar as the decision is an important one. Unimportant hard cases cannot be characterized as constituting intransitive reasoning.

[13] Whether this means that hard cases should rather be identified purely based on the first two premises of the small-improvement argument, as we have suggested in this paper, is a question for future research.

Appendix: Vignettes of Study 1A and Study 2A

Study 1A

First Vignette

Drink Choice 1

Some people think that a cup of tea is neither better nor worse than a cup of coffee. Others believe that a Gin & Tonic is neither better nor worse than a Bloody Mary. Imagine you are sitting in a café which only has two different drinks on the menu. Please name two drinks for which YOU could not tell which one of those is better or worse than the other.

- First drink:. . .
- Second drink:. . .

Event Choice 1

Some people think that a Bryan Adams concert is neither better nor worse than attending the Super Bowl. Others believe that going to a Broadway musical is neither better nor worse than seeing the Chinese National Circus. Imagine it is Saturday night, and you have got the choice only between two cultural or sports events. Please name two events for which YOU could not tell which one of those is better or worse than the other.

- First Event:. . .
- Second Event:. . .

Career Choice 1

Some people think that studying art is neither better nor worse than studying law. Others believe that studying philosophy is neither better nor worse than studying economics. Imagine you have got the choice only between studying two different courses. Please name two courses for which YOU could not tell which one of those is better or worse than the other for studying.

- First course:. . .
- Second course:. . .

Second Vignette

Drink Choice 2

Now, imagine that the waiter tells you that you will get a discount of 5% on one of these drinks. Keeping in mind the two drinks you selected before, how would you evaluate your situation?

- The drink plus a small discount is **better** than the other drink without a discount.
- The drink plus a small discount is **worse** than the other drink without a discount.
- The drink plus a small discount is **neither better nor worse** than the other drink without a discount.

Event Choice 2

Now, imagine that the ticket salesperson tells you that you will get a discount of 5% on one of these events. Keeping in mind the two events you selected before, how would you evaluate your situation?

- The event plus a small discount is **better** than the other event without a discount.
- The event plus a small discount is **worse** than the other event without a discount.
- The event plus a small discount is **neither better nor worse** than the other event without a discount.

Career Choice 2

Now, imagine that the university tells you that you will get a discount of 5% on the costs of one of these courses of study. Keeping in mind the two courses you selected before, how would you evaluate your situation?

- The course plus a small discount is **better** than the other course without a discount.
- The course plus a small discount is **worse** than the other course without a discount.
- The course plus a small discount is **neither better nor worse** than the other course without a discount.

Third Vignette

Drink Choice 3

Now, lastly, imagine you have got the choice between the drink for which you were offered a 5% discount or the very same drink without getting the discount.

How would you evaluate your situation?

- The drink plus a small discount is **better** than the very same drink without a discount.
- The drink plus a small discount is **worse** than the very same drink without a discount.
- The drink plus a small discount is **neither better nor worse** than the very same drink without a discount.

Event Choice 3

Now, lastly, imagine you have got the choice between the event for which you were offered a 5% discount or the very same event without getting the discount.

How would you evaluate your situation?

- The event plus a small discount is **better** than the very same event without a discount.
- The event plus a small discount is **worse** than the very same event without a discount.
- The event plus a small discount is **neither better nor worse** than the very same event without a discount.

Career Choice 3

Now, lastly, imagine you have got the choice between the course of study for which you were offered a 5% discount or the very same course without getting the discount.

How would you evaluate your situation?

- The course plus a small discount is **better** than the very same course without a discount.
- The course plus a small discount is **worse** than the very same course without a discount.
- The course plus a small discount is **neither better nor worse** than the very same course without a discount.

Study 2A

The first two vignettes were identical to Study 1. We therefore only list the third vignette of Study 2.

Drink Choice 3

In the beginning, we asked you to imagine two drinks for which you could not tell which one of those drinks is better or worse than the other.

Would flipping a coin be an appropriate way to decide in favor of one of the options?

- Yes
- No

Please briefly explain your answer:...

Event Choice 3

In the beginning, we asked you to imagine two events for which you could not tell which one of those events is better or worse than the other.

Would flipping a coin be an appropriate way to decide in favor of one of the options?

- Yes
- No

Please briefly explain your answer:...

Study Choice 3

In the beginning, we asked you to imagine two courses of study for which you could not tell which one of those courses is better or worse than the other.

Would flipping a coin be an appropriate way to decide in favor of one of the options?

- Yes
- No

Please briefly explain your answer:...

References

Andreou, C. (2011). Choosing Well: Value Pluralism and Patterns of Choice. In T. Brooks (Ed.), *New Waves in Ethics* (pp.48-63). Basingstoke: Palgrave Macmillan.

Andreou, C. (2015). Parity, Comparability, and Choice, The Journal of Philosophy CXII, 5-22.

Boot, M. (2007). *Incommensurability, Incomplete Comparability and the Scales of Justice*. Dissertation (University of Oxford).

Boot, M. (2009). Parity, Incomparability and Rationally Justified Choice, *Philosophical Studies*, 146 (1), 75-92.

Bridges, D. & Mehta, G. (2013). *Representations of Preferences Orderings*. Berlin: Springer.

Chang, R. (2002a). The Possibility of Parity. *Ethics*, 112 (4), 659-688.

Chang, R. (2002b). *Making Comparisons Count*. New York: Routledge.

Chang, R. (2012). Are Hard Choices Cases of Incomparability?. *Philosophical Issues* 22 (1), 106-126.

Chang, R. (2013). Commitment, Reasons, and the Will. In *Oxford Studies in Metaethics* 8, 74-113.

Chang, R. (2015). Transformative Choices. *Res Philosophica*, 92 (2), 237-282.

Chang, R. (2014). The existentialist of hard choices (Interview by Richard Marshall with Ruth Chang) *http://www.3ammagazine.com/3am/the-existentialist-of-hard-choices/*

De Silva, I. (1991). Consistency Between Reproductive Preferences and Behavior: The Sri Lankan Experience. *Studies in Family Planning*, 22 (3), 188-197.

De Sousa, R. (1974). The Good and the True. *Mind*, 83 (332), 534-551.

Espinoza, N. (2008). The Small Improvement Argument. *Synthese*, 165 (1), 127-139.

Fishburn, P (1991). Nontransitive preferences in decision theory. *Journal of Risk and Uncertainty*, 4 (2), 113-134

Gustafsson, J.& Espinoza, N. (2010). Conflicting Reasons in the Small-Improvement Argument. *The Philosophical Quarterly*, 60(241), 754-763.

Hsieh, N. (2005). Equality, Clumpiness and Incomparability. *Utilitas*, 17 (2), 180-204.

Lambooij, M. & Harmsen I. & Veldwijk, J. & de Melker, H. & Mollema, L. & van Weert, Y. & de Wit, A. (2015). Consistency between stated and revealed preferences: a discrete choice experiment and a behavioural experiment on vaccination behaviour compared. *BMC Medical Research Methodology*, 15 (19), 1-8.

Lindman H. & Lyons J. (1978). Stimulus complexity and choice inconsistency among gambles. *Organizational Behavior and Human Performance* 21 (2), 146Ű159.

Loomes, G. & Taylor, C. (1992). Non-Transitive Preferences Over Gains and Losses. *The Economic Journal*, 102 (411), 357-365.

Loureiro, M. & McCluskey, J. & Mittelhammer, R. (2003). Are Stated Preferences Good Predictors of Market Behavior?. *Land Economics*, 79 (1), 44-55.

Messerli, M. & Reuter, K. (2016). Hard Cases of Comparison. *Philosophical Studies*, 10.1007/s11098-016-0796-y

Parfit, D. (1986). *Reasons and Persons*. Oxford: Clarendon Press.

Peterson, M. (2013). The Dimensions of Consequentialism. Cambridge: Cambridge University Press.

Raz, J. (1986). *The Morality of Freedom*. Oxford: Clarendon Press.

Schaber, P. (1994). Sind alle Werte vergleichbar? Kosten-Nutzen-Analyse und das Inkommensurabilitätsproblem. *Analyse und Kritik*, 16 (2), 153-165.

Tversky, A. (1969). Intransitivity of preferences. *Psychological Review*, 76, 31-48.

Ulmann-Margalit, E. & Morgenbesser, S. (1977). Picking and Choosing. *Social Research*, 44 (4), 757-785.